Volume 9 • Number 3 • 2006
VALUE IN HEALTH

Bringing Health Economic Modeling to the 21st Century

David Eddy, PhD, MD

Archimedes, Inc., Oakland, CA, USA

This is a summary of the presentation given by David Eddy at the ISPOR 10th Annual International Meeting Second Plenary Session, May 17, 2005, Washington, DC, USA. Dr Eddy is Founder and Medical Director of Archimedes, Inc., Oakland, CA, USA.

At the core of every health economic analysis is a model. It is the model that converts data about diseases, interventions, and costs into projections of what will happen in the future. The importance of good data is emphasized by the expression "Garbage in, garbage out." With a flawed model "Anything in, garbage out." My purpose in this commentary is to argue that the increasing complexities of the problems we are trying to analyze are pushing our current models to their limits, and to offer a new type of model for our toolkit. Because both the need for the model and its design are rooted in my experiences, I will tell the story from a personal perspective.

Need for a New Type of Model

The need for a new type of model was gradually forced on me by the gap between the types of problems I wanted to analyze and the capabilities of the existing models. My own experiences began about 35 years ago with Markov models [1]. Given the continuing prominence of this type of model in health economic modeling, this is a reasonable place to start.

Every reader of *Value in Health* is familiar with the basic structure of a Markov model. The fundamental property that makes a random process a Markov process is that given the present, the future is conditionally independent of the past. Other ways to say this are that history does not matter, and the process has "no memory." Typically a Markov process is defined in terms of discrete states; at any time the process will be in a particular state, and at discrete time intervals the process can make transitions between states. A textbook example is cars in a queue for a tollbooth. The state of the process is the number of cars in the queue at any time. The state changes every time a new car enters the queue or the front car is cleared through the tollbooth, and the trajectory of the process is determined by the probabilities of those events.

In a typical health economic model the process is a pathological condition such as a disease, and the states are constructed by dividing the pathological process into discrete stages. An example is to represent cardiovascular disease (CVD) by four states: "no CVD," "had an MI," "angina," and "dead of CVD." Another is to represent nephropathy as "no nephropathy," "microalbuminurea," "proteinuria," "end-stage renal disease (ESRD)," and "dead of ESRD." Progression from one state to another is determined by the transition probabilities, with transitions typically occurring at annual intervals.

Markov models have some very useful qualities, especially when compared with other types of models such as regression equations and decision trees. The main strength that distinguishes them from those other models and that makes them so popular for health economic modeling is that they are dynamic and probabilistic—they can address problems in which events and decisions are occurring, subject to chance, over time. Other strengths are that the concept of discrete states resonates with the labels we often use to talk about diseases, such as a person having "diabetes" or having "hypertension." The notion that diseases progress from state to state is also intuitively appealing, as exemplified by images of cancers progressing from in situ, to localized, to regional metastases, and so forth. The annual interval corresponds to many other things that are tallied annually, such as annual incidence rates and annual budgets. States and transitions are useful hooks on which to attach costs and quality weights. And the basic mathematical structure of a Markov model is easy to learn, describe, and calculate. In short, if the states can be defined so that they are sufficiently realistic for the problem to be analyzed, if there are good data for the transition probabilities, and if the Markov history-does-not-matter assumption holds, then the Markov model can be a very useful framework for calculating clinical and economic events.

Unfortunately, the three conditions just listed are difficult to meet. The reason is that biological processes and clinical medicine are fundamentally different than the types of processes for which Andrei Markov designed his model. As a consequence all sorts of sim-

10.1111/j.1524-4733.2006.00097.x

1098-3015/06/168 168–178

plifications and assumptions have to be made to fit them into that framework. Consider first the fundamental Markov assumption that history does not matter. Everyone knows that this is patently untrue. Imagine opening a patient's chart and seeing only today's notes (e.g., today's physical findings, today's lab results)—all the previous pages have been discarded. The very nature of a chronic condition is that the risk of complications, such as stenosis in arteries, builds up over time. But the no-history assumption requires, for example, that a person's cholesterol-related risk of a myocardial infarction (MI) depends only on their current low-density lipoprotein (LDL) level, say, 110 mg/dl, and it does not matter whether they have always been at 110 mg/dl or they have been at 175 mg/dl for 30 years and were just brought down to 110 mg/dl by treatment. The same point can be made about blood pressure, glucose, smoking and other risk factors that lead to chronic damage. Thus, the no-history assumption creates a problem. Either we have to limit ourselves to problems where history truly does not matter, which is a small list in medicine; or we have to pretend that it does not matter and accept whatever inaccuracies that may cause; or we have to find some way to overcome the assumption, which is in essence saying that we do not like the fundamental piece of the Markov framework. Incidentally, the same assumption of no-history is being made every time we use an equation like the Framingham and insert current values for cholesterols, blood pressure and smoking, or change the values to estimate the effects of treatments.

Now consider the concept of a discrete state. To appreciate the problems here we first need to recognize that because the number of possible transitions grows exponentially with the number of states, and because data have to be found for every transition, there is a very strong motivation to keep the number of states down. So although in theory the Markov framework can handle a very large number of states, in practice they are kept to a small number, typically under a dozen. But reducing the number of states means that each state has to include more disparate groups of people. This can happen in two main ways—either by dividing a disease process into a small number of discrete states, or by combining several disease processes into a single state. Take the first one first. An immediate observation is that as comfortable as we might be talking about diseases using clean labels like "heart attack," "diabetes," "ESRD," and "regional metastases," these are enormously complex conditions and they do not jump from one discrete state to another. A heart attack is not a single variable that switches suddenly from "no MI" to "MI," like a new car entering a queue; it is the result of a multidimensional process that progresses continuously and is affected by a wide variety of factors most of which themselves are continuously changing. A modeler who dichotomizes that process into "no MI" and "MI" does not do so because he or she believes it is accurate, but because the discrete-state structure of a Markov model requires this type of simplification.

The problem of course is that the simplifications hide a huge number of important factors that can have profound effects on the outcomes, but that get ignored in the calculations; imagine condensing an MI patient's chart down to one sentence that says "Has not had a MI." A trivial example is that a person with 90% occlusion of the left anterior descending artery and a person with clean arteries would both be in the same state of "no MI," and a model that recognizes only "no MI" and "MI" will treat them as though they were the same. We might try to fix the problem by stratifying the state in some way. But even if we look at populations that have similar descriptions there can be wide differences in future risks that the model cannot see because of the simplified states. For example, let's narrow the state to "no MI but high risk of CAD"; the 5-year risk of MIs can still vary from 4% [2] to 6% [3] to 8% [4] to 13% [5,6]. Consider the state "had an MI"; the 5-year risk of a repeat MI in people with a previous MI can vary from 13% [7,8] to 25% [9] depending on other variables that are present in the real population but are missed in the designation "had a MI." Needless to say differences like these, by factors of 2 or 3, have disastrous consequences for an analysis.

To try to sharpen the definitions of the "no MI" and "MI" states even further we might turn to the Framingham equation, but that maneuver only tightens the argument I want to make, which is that the basic Markov framework is not working for us. First, the Framingham equation only includes a handful of variables. For example it does not include such obvious factors as family history, presence of angina, use of aspirin, and the duration, severity or treatment of diabetes. Thus, although it may help solve part of the problem, it does not solve all of it. Second, even acknowledging its limitations we do not have equations like the Framingham for the overwhelming majority of other conditions. Third, as already described the Framingham equation itself makes a no-history assumption. But the fourth and most important point is that pulling in the Framingham equation to save the "no MI" state is a clear statement that the "no MI" state is too simplistic. The "no MI" state has become a mere label for a myriad of states defined on the fly by the independent variables in the Framingham equation. It is not a Markov state any more; its only real role is to serve as a storage point for tallying events.

The second way the number of discrete states is kept down is by combining multiple disease processes into a single state. For example, people who: 1) "have had two or more depressive symptoms for at least two weeks, and have functional impairment, but do not

meet the *Diagnostic and Statistical Manual of Mental Disorders* criteria for major depression"; 2) "have minor depression"; 3) "have dysthymia"; or 4) "have major depression in partial remission" may be combined into a state called "significant depressive symptoms." Every clinician knows that these four clinical conditions are in fact very different, which is why they have different names and diagnostic criteria, as well as different etiologies, natural histories, symptoms, treatments, and prognoses. But by combining them into a single state, we are telling the Markov model to assume they are the same.

The transition probabilities also present problems. In medicine they are not simple chance events; they are the result of very complex biological phenomena and our attempts to manage them. Indeed almost everything that is interesting in medicine is happening inside the transitions. Consider the factors that determine whether a person jumps from "no MI" to "had an MI" or "dead from MI." These probabilities are determined by risk factors such as dyslipidemia, hypertension, diabetes, obesity, and smoking; interventions to manage the risk factors; use of medications like aspirin; how alert a person is to early signs and symptoms; the availability of emergency systems and delays in getting care; the role of symptoms and tests like the EKG's, CK-MB or Troponin in making a diagnosis; protocols for deciding who to admit, observe or send home; emergency treatments like aspirin, thrombolytics, percutaneous transluminal coronary angioplasty (PTCAs), and stents; and longer-term interventions like cholesterol lowering drugs, beta blockers, and rehabilitation.

The importance of this complexity for building a Markov model is profound. First, the fact that the transition probabilities are affected by so many factors greatly complicates the task of finding data for the transition probabilities, essentially requiring that there be data from other settings that match the setting we want to address (e.g., the same population, same management of risk factors, same use of aspirin, same triage protocols, same acute care, etc.) If even one of these factors is off, the transition probability and therefore the final results will be off. For an example consider that to get the rate of repeat MIs in people with previous MIs one modeler might go to the CARE trial and come back with a 5-year rate of 13% whereas another might go to the 4-S trial and come back with 25%. Second, if we can not find the data we need from a matching setting, or if the problem we want to address involves any of the things that are inside the transition probabilities, we will need to build another model (and find data for it) just to calculate the transition probability in the Markov model. For example, suppose the problem is to evaluate the use of troponin in the work-up of people who present with chest pain in the ER. To analyze this problem we would need to build another model and find data for the transition probabilities from "no MI" to "had an MI" or "died of MI" for each of the possible roles for troponin we wanted to explore (e.g., Use troponin instead of CK-MB? In addition to CK-MB? Use troponin I? Or troponin N? Use a cutoff point of ≥0.4 ng/ml? Or ≥0.7 ng/ml?)

Finally, the use of discrete time intervals, usually annual, also creates problems. In fact, almost nothing real happens at annual intervals. Important clinical events can occur over minutes or hours, as in getting cardiopulmonary resuscitation or thrombolytics to a person with a heart attack, or over years, as in a healthy 30-year-old. A chronic disease like diabetes can plod along for years with no clinical symptoms, and then erupt with acute complications. Trying to capture these effects in a discrete time interval puts even more burden on the transition probabilities and states.

At this point it is very important to stress that these limitations do not necessarily invalidate a Markov model. But as already indicated for the no-history assumption, they do mean one of three things: either 1) we have to be willing to ignore the importance of these simplifications and assumptions, and accept whatever inaccuracies they may create; 2) we have to restrict ourselves to applications that are truly not affected by the limitations (e.g., it truly does not matter that we are lumping together all people with MIs or all people with "significant depressive symptoms"); or 3) we have to build additional models to address each of the important factors, essentially replacing the transition probabilities and the states with other models. The first is unacceptable and the second is giving up on a large class of problems. The third is the only acceptable way to use a Markov framework to answer important questions. But it leads to other problems.

To illustrate these problems I will pick on my own work. Around 30 years ago the American Cancer Society asked me to help them update their recommendations for cancer screening. I built a Markov model to calculate the costs and effectiveness of different screening strategies [10,11]. The model was successful in the sense that its results caused major changes in the recommendations [12], some of which are still reverberating today (e.g., the 3-year Pap smear, 3- to 5-year colon cancer screening, no routine sputum cytology or chest x-rays even for smokers, mammography only for women over age 50). The analysis introduced concepts of modeling, evidence, and cost-effectiveness into national guidelines [11]. The resulting guidelines attracted considerable public attention [13]. The predictions made by the model were validated against an independently collected data set 6 years later [14].

So what's the problem? The problem is that extraordinary mathematical contortions were needed to stuff the cancer-screening problem into the Markov framework. We have already seen that the Markov

framework involves discrete states and assumes no memory. In contrast, in the cancer-screening problem virtually everything changes continuously and nothing obeys the Markov no-memory assumption. Cancers are continuously changing over time; there are multiple tests that can be carried out in any order or frequency—and not necessarily together or annually; the sensitivities of the screening tests vary depending on how far the cancers have advanced (e.g., the size of a breast mass); the cancers can become apparent between scheduled screening tests through self examination or symptoms; the prevalence of detectable but not yet detected cancers depends on the past use of the tests and patient self detection; how early the cancers are detected also depends on those factors; the effectiveness and costs of treatments depend on how early the cancers are detected, and so forth. To accommodate all these factors within the Markov structure I had to build sub models that redefined every state and recalculated every transition probability on the fly. It required dozens of differential equations, and a 250-page book to describe it all [15]. Although the final model was a Markov model in the sense that it used states and transitions, all the heavy lifting was done by the differential equations whirring in between every transition. The amount of new mathematics needed to get around the limitations of the Markov structure is indicated by the fact that the model won a prize for the most important contribution in the English language to the fields of operations research and management science [16].

The point is that somewhere down the line one has to stop and ask oneself, "If I have to go through so many gyrations to fit a problem into a Markov framework, am I really using the right model? Maybe there is some other mathematical formulation that is a better fit to biological and clinical problems."

Design of a New Type of Model

And of course there is. It is calculus, developed more than 300 years ago by either Isaac Newton or Gottlieb Leibniz depending on which side of the Channel you're on. Indeed, if one looks at my screening model with a fresh eye it is obvious that it is really a set of differential equations and the Markov structure is little more than a framework for tabulating the events and costs. I was not using the Markov properties—no history, discrete states, or discrete time—at all. In fact I was doing everything I could to get around them. Furthermore, by breaking up the progression of the cancers into artificial states the Markov structure was creating unnecessary and damaging constraints. Why not just shuck off the Markovian constraints and simply build the model directly in differential equations?

This theme kept recurring in my subsequent work on guidelines, coverage policies, performance measures, quality improvement, cost-effectiveness, and related topics. Over time, it became more and more apparent that, at least for the problems that I was trying to address, a new type of model was needed.

Gradually, the following motivations and design criteria emerged. First, to help clinicians make decisions; to help design guidelines, performance measures, and the "what-to-do" parts of disease management programs; and to be credible, the model had to start at the level of physiological and clinical detail at which clinicians think. Essentially, it had to encompass all the biological variables that physicians consider to be important in the management of their patients. This level of detail would also be required to help analyze the physiological processes underlying diseases and their treatments, and to help design, interpret, and extend clinical trials. Second to address issues that arise in the design of the "how-to-do-it" parts of disease management programs, case management protocols, and continuous quality improvement projects, the model had to include care processes, logistics, and behaviors at an equally high level of detail. This is also needed to translate idealized clinical trial results into realistic settings. Third, to provide credible information about logistics and cost effectiveness, it had to include system resources such as facilities, personnel, visits, admissions, equipment, and all their costs. It should be able to track events and their costs just as is done in real health-care systems. Fourth, to help set clinical priorities, design strategic goals, and prioritize and/or combine performance measures, the model had to be able to span across all types of interventions (primary prevention, screening, diagnosis, treatment, secondary prevention, and support care) and span multiple diseases using the same methodology. A broad span would also be required to address patients who have multiple diseases (comorbidities), syndromes that affected multiple organ systems, drugs that have multiple effects, and combinations of drugs. Fifth, to address questions of timing—such as screening, frequency of follow-up visits, or how long a medication should be tried before the dose is changed—the model had to function in continuous time, and be able to address events that can change as rapidly as minute by minute, or as slowly as years. Finally to be credible, the model had to be able to simulate most important epidemiological studies and clinical trials at the level of clinical detail at which they are designed and reported, and match or predict their results within the appropriate sampling errors.

It was already clear that differential equations are the natural way to represent the continuous changes and interactions of biological variables. The next question was whether there are better ways to represent all the other aspects of a health-care system such as protocols, care processes, providers, visits, resources, and costs. And again, there are. One is object-oriented pro-

gramming (OOP). Although much younger than calculus—about 35 years old—it is very powerful and versatile, being the programming method behind such disparate things as graphical user interfaces, Excel, airline reservation systems, video games, and Defense Department models of the European war theater. It can handle very high levels of complexity and detail, although remaining flexible and easy to update. It was an obvious choice for a framework for health economic modeling.

The Archimedes Model

The use of differential equations and OOP can be illustrated with the Archimedes model, which a team led by Len Schlessinger and me has been developing over the last 12 years. The mathematical formulation and a clinical overview of the model have been described elsewhere [17,18]. Examples of specific equations and sources are available through our Web site (http://www.archimedesmodel.com). Briefly, the Archimedes model uses differential equations, OOP, and a modeling concept we call "features" to represent human physiology at a level of detail roughly comparable to that found in general medical textbooks or patient charts. For example, for analyzing something like the metabolic syndrome the model includes variables relating to glucose metabolism (hepatic glucose production, uptake of glucose by fat and muscle, insulin amount, insulin resistance, fasting plasma glucose [FPG], HbA1c, and 2-h OGT); lipids (e.g., LDL, high-density lipoprotein [HDL], triglycerides, LDL particle size, and number); obesity (body mass index [BMI] and waist circumference); inflammation (e.g., C-reactive protein); blood pressure (e.g., cardiac output, arterial compliance, peripheral resistance); coronary artery disease (e.g., gradual stenosis, plaque rupture, myocardial ischemia); strokes (hemorrhagic and ischemic); as well as models of nephropathy, retinopathy, and neuropathy. Time is continuous, all the biological variables are continuous functions of time, and any event can occur at any time. The model includes multiple organ systems and diseases as part of a single physiology. It also includes a detailed representation of other part of the healthcare system, as described in the design criteria listed in a previous paragraph.

Our objective is to create a virtual world that can be used for such things as exploring the effects of guidelines, disease management programs, performance measures, and quality improvement programs; designing, predicting, and interpreting clinical trials; forecasting clinical and economic outcomes; setting priorities; and estimating person-specific outcomes. The scope of potential applications is most easily described by listing some of the actual applications in the past year.

- "How does treating insulin resistance directly compare with treating hyperglycemia? (Insulin resistance affects not only glucose but also triglycerides, HDL, BP and other variables.);
- "What would be the effect on biological variables and clinical outcomes of a drug that decreases weight by x%? For what populations would such a drug be best indicated?
- "How does delivering insulin through inhalation compare with the current method of injection? What are the implications of the different effects on HbA1c for downstream clinical outcomes? What trial should we conduct to determine the appropriate indications?
- "Is there an underlying cause of the metabolic syndrome? What are the roles of insulin resistance, inflammation, adiponectin, LDL particle size, apolipoprotein B, and fibrinogen? What proportion of CVD events can be attributed to the metabolic syndrome compared with nonmetabolic causes such as sex, age, race/ethnicity, and smoking? How should the metabolic syndrome be defined?
- "We want a calculator that will tell people their risks of diabetes and its complications, taking into account not only the usual Framingham type variables (sex, age, SBP, TC/HDL, smoke? diabetes?), but also duration and severity of diabetes, past medical history (e.g., previous MI), past treatments, past and current weight/BMI, current symptoms and complications, and current medications.
- "What are the relative effects of raising performance from the average to the 90th percentile level for each of the HEDIS measures for CVD, diabetes, CHF and tobacco? If we can only focus on five measures, which are the most important?
- "We've done a phase II trial that showed these results . . . What would a phase III trial show for clinical outcomes? What would happen in different populations (i.e., different indications)? What is the optimal design for a trial?
- "We want to combine two drugs, but we are not willing to assume the effects are additive or multiplicative. What will be its effects in populations that have different initial risks, different current treatments, and different current degrees of control?
- "There is a trial in progress that will be reported in the AHA meeting in two months. We want you to predict its results so we can plan."

A frequently asked question about the Archimedes model is whether the high level of physiological detail is really required. The answer depends on what types of questions you want the model to be able to address. For us the answer is in the projects just listed. If we

want to be able to address the metabolic syndrome, then the model has to include insulin resistance, adiponectin, C reactive protein, etc. If we want to study the cost-effectiveness of various strategies for meeting Health Employer Data Information Set (HEDIS) targets, then the model has to include the HEDIS variables, their tests, and treatments, and all the logistics that will be affected. It also has to include all the pertinent conditions (e.g., CVD, diabetes, congestive heart failure [CHF]) in a single integrated model. If you want to analyze a disease management program for CHF that uses an ejection fraction <50% as one of the criteria for selecting patients, the model has to include ejection fractions and all the physiology around it. Our criterion is that if the people who know the field and will use the model believe a variable is important, and if the data support them, we want to include it. Stated another way, we want the details of the questions to determine the details of the model, not the other way around.

Validation of the Model

Archimedes illustrates the type of model that can be built with differential equations, OOP, and features. But it also raises a question. How can we determine whether it works? That is, how can we know whether its results are accurate in predicting what will actually occur? To answer this we need to go back to the scientific method, the foundation of which is observation of real events. We need to use the model to simulate actual events that occur in the real world, and then compare the model's results to the real results.

To test the Archimedes model the real results we have chosen to compare ourselves to are clinical trials [19], although we also use epidemiological studies and smaller studies to check parts of a model. We chose clinical trials for the same reasons they are so valuable in other aspects of medicine; they are the most rigorous and best-documented way to determine what is really happening. Clinical trials are the foundation on which the clinical practice of medicine it built, and by simulating them we are helping ensure that the model anchored just as strongly to that foundation. Methodologically, simulating clinical trials also has the virtue of testing an entire model—from the characteristics, behaviors, histories, and past treatments of the population; through the occurrence of symptoms, visits, tests, diagnoses, and treatments; and finally to the occurrence of the clinical outcomes.

Briefly, the steps are: 1) use the inclusion and exclusion criteria as well as information on the distribution of characteristics, biological variables, current and past medical histories, medications, and behaviors (what is often in "Table 1" of published reports of trials) to select a simulated population that matches the real population; 2) give this population interventions according to the protocols of the trial; 3) run the model; 4) measure the outcomes that define the endpoints of the trial, using the definitions and measurement protocols specified for the trial; and 5) compare the model's results to the results seen in the real trial. Although our main focus is on the primary and secondary endpoints defined for each trial, we also check the model's accuracy for calculating biological outcomes. All of the simulations are performed at the highest level of detail of which the model is capable. For example, if two trials reported retinopathy outcomes but one measured two-step retinopathy [20], whereas the other measured three-step retinopathy [21], we have the simulated physicians do simulated eye exams on each simulated patient using the particular protocol that applies to each trial.

To avoid obvious selection and reporting biases we have an independent committee pick the trials and monitor the results; for the first round of validations the American Diabetes Association set up the committee. The trials are the major ones that pertain to the questions the model will be used to address; in this case, those relating to diabetes and its complications. The validations are performed with the same version of the model, as the results would be meaningless if the model were revised each time to make it fit each trial. We try to find enough trials should so that they "triangulate" the problems for which the model is to be applied; we want to validate the model against at least one trial that involves the same population, at least one that involves the same treatments, and at least one that involves the same types of outcomes. And we want as many of the trials as possible to be independent in that they were not used to build the model. For the first round, 40 of the 74 validation exercises were independent.

Examples of results for the Archimedes model are shown in Figures 1–6. Figures 1 and 2 evaluate the model's ability to reproduce biological outcomes, in this case the average FPG levels. Figure 1 illustrates the results for the UK Prospective Diabetes Study (UKPDS) trial of conservative versus intensive management of glucose in people with newly diagnosed diabetes. Because we used some information from the UKPDS trial to help build the Archimedes model, this is a dependent validation. An example of an independent validation of biological outcomes is the Diabetes Preventions Program (DPP), which evaluated the effects of placebo versus metformin versus intensive lifestyle modification in people with prediabetes [22] (Fig. 2). Figure 3 illustrates the calculation of a clinical outcome, the rate of coronary artery events in the UKPDS (another dependent validation). An example of an independent validation for a clinical outcome is the Heart Protection Study [5] in which adults at high risk of MI because of high LDL, other arterial occlusive disease or diabetes were treated with simvastatin or a

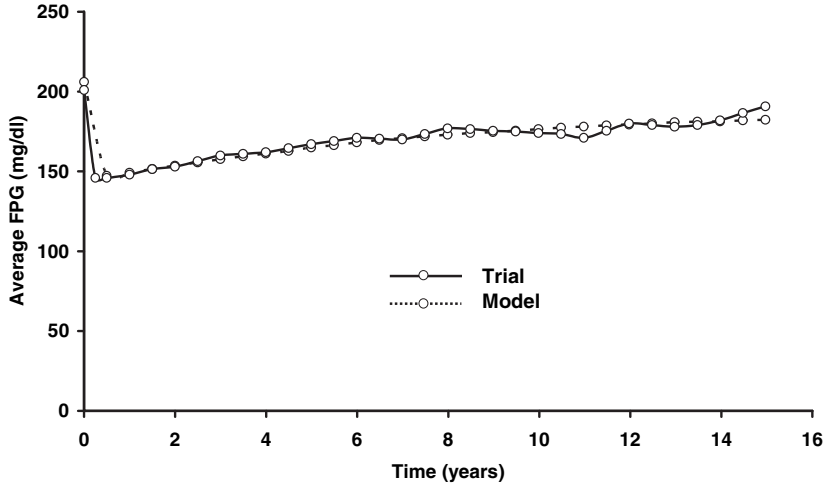

Figure 1 Fasting plasma glucose in the control group of UK Prospective Diabetes Study: comparison of trial and model. FPG, fasting plasma glucose.

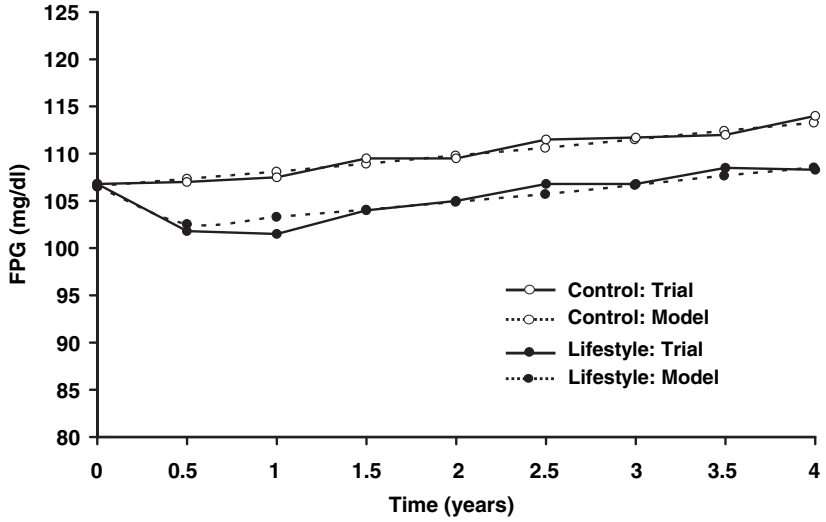

Figure 2 Fasting plasma glucose in the Diabetes Preventions Program trial.

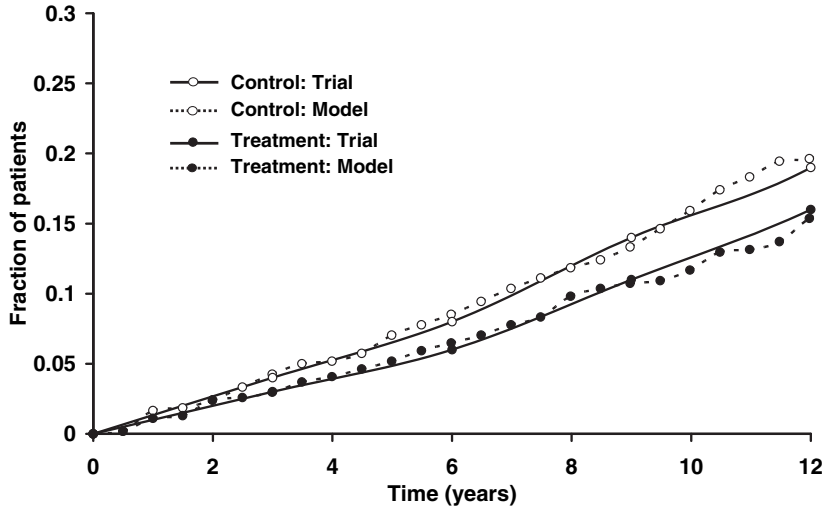

Figure 3 UK Prospective Diabetes Study: myocardial infarction (fatal and nonfatal).

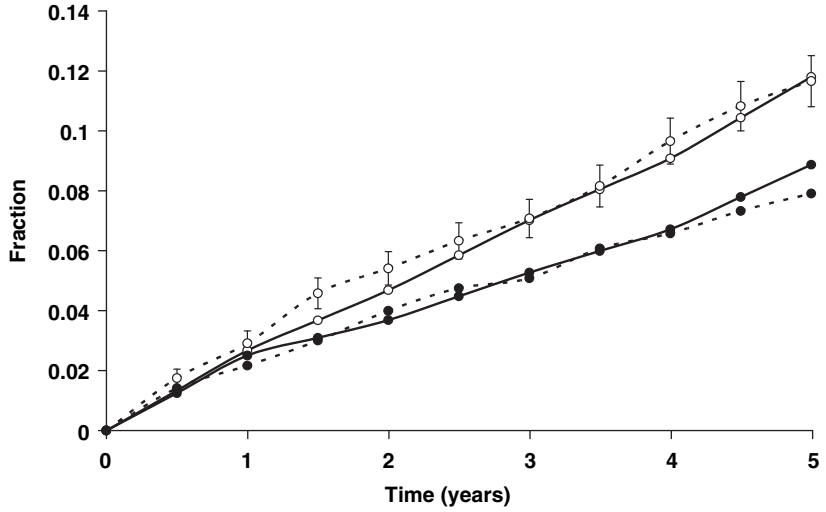

Figure 4 Major coronary events in the Heart Protection Study.

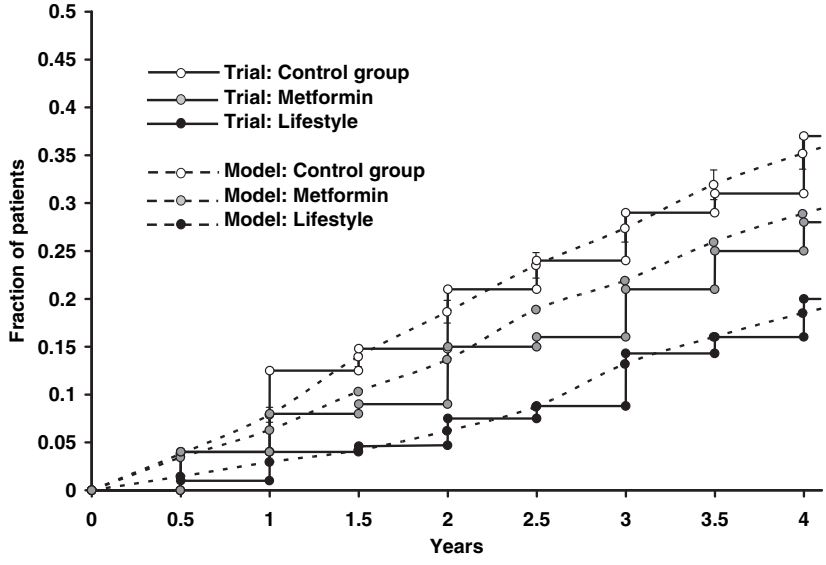

Figure 5 Archimedes prediction of Diabetes Prevention Program results.

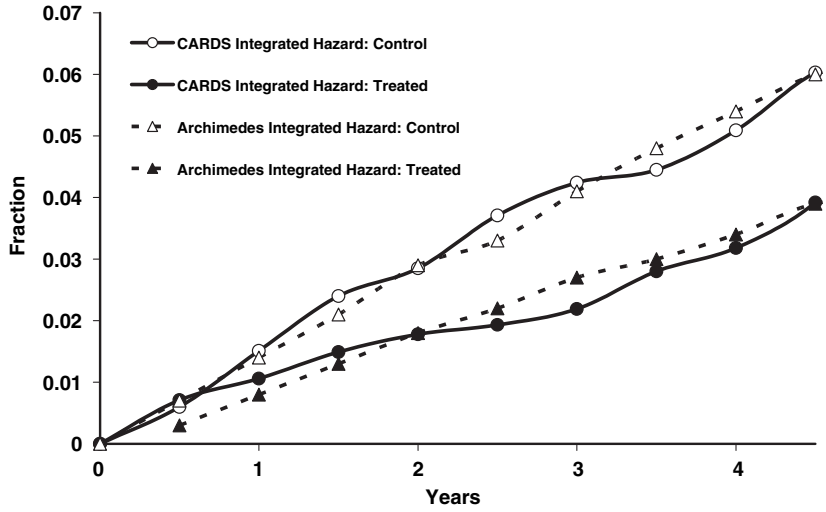

Figure 6 Archimedes prediction of Collaborative Atorvastatin Diabetes Study trial: major coronary events.

placebo (Fig. 4). When possible we try to predict the results of trials before their results are known or published. Two examples are the DPP (Fig. 5) and the Collaborative Atorvastatin Diabetes Study trial [23] that compared atorvastatin versus placebo in people with diabetes (Fig. 6). Results for 15 other trials are described elsewhere [19]. Overall, the correlation between the model's results and the real results is about 0.98. Our practice is to continue to test the model against new trials as they appear, or if they become important benchmarks for new analyses. Thus, far the Archimedes model has been validated against 28 trials.

Validations like these do not promise that every new analysis or prediction with the model will be perfectly accurate, but they do at least demonstrate that a model built at a high level of detail using differential equations, OOP, and features can be reasonably accurate in representing what is currently known from existing trials. The relatively high success rate with more than 40 independent validation exercises ($r^2 = 0.96$) also suggests that existing information, at least as it is represented in the Archimedes model, is reasonably complete in enabling predictions of new trials the model has never seen. This is encouraging, because clinical trials are very analogous to the types of problems we are building the model to address; like real problems they involve specific populations given specific interventions using specific protocols to affect specific results. It is also encouraging for medicine in general.

We need to emphasize that there will always be mismatches and surprises with any model. That should not upset anyone. Indeed, the whole purpose of conducting new clinical trials is to find results that cannot be predicted with existing information. In our experience, there have been two occasions out of more than 40 independent validation exercises when the model's results did not "statistically match" the real results: the rate of MIs in the control group of the WOSCOPS trial of pravastatin versus placebo in high-risk men in western Scotland, and the effect of atorvastatin on the rate of strokes in people with diabetes. When this occurs, we determine the cause and if necessary use the new information to improve the model. We then look for other studies to independently validate the modifications. Continuous application of this process gradually expands and improves the model.

Conclusions

Different models have different strengths and weaknesses. The task of the modeler is to find the best type of model to match the characteristics of the system to be represented and the problem to be solved. Regression models are excellent methods for mining data sets and identifying relationships between variables. Decision trees are powerful ways to select between options to maximize expected value and take into account risk aversion. Similarly, Markov models can be powerful tools for representing systems that are dynamic and probabilistic. The fact that many health economic problems are also dynamic and probabilistic makes it an obvious choice for this field.

But each type of model also has limitations. Here the task of the modeler is to understand the limitations and either restrict the application of the model to problems that stay within the limitations, or modify the model to get around the limitations. An unfortunate fact about Markov models is that several of its fundamental properties—a process whose future is conditionally independent of the past, characterization of a process as discrete states, representation of movements between states as simple transition probabilities, and cutting time into discrete intervals—are not a good match to biological systems and clinical problems. Mathematicians have been ingenious in developing methods to expand the capabilities of the Markov framework, with such things as semi-Markov continuous, partially observable and hidden Markov processes. For at least 30 years people have been adding decision nodes, bridge models, and parallel models to address the special needs of health care. But as we try to push the Markov model farther and farther, the simplifications and assumptions we have to make become more and more troubling, the modifications we have to make to get around the simplifications and assumptions become more and more difficult, and the results become more and more suspect. At some point, it is appropriate to ask whether there is not a better way.

In our experience there is a class of important problems that push the Markov framework beyond its intended limits. Indications that the limits are being reached are problems that involve one or more of the following:

- Diseases that are multifactorial and not well represented as a single process. Diseases that are too complex to be well described by a small number of discrete states.
- Diseases in which past medical history is important.
- The need to represent the underlying biology of a condition or mechanism of action of a drug on biological outcomes.
- Clinical processes such as guidelines, disease management, and quality improvement programs where the steps in the process have implications for effectiveness, costs or logistics.
- Comorbidities and syndromes that affect multiple organ systems and outcomes.
- Drugs with multiple effects and combinations of drugs.

- Cost effectiveness problems in which it is important to include logistics such as tests, treatments, visits, admissions, etc.
- Problems where patient and/or physician behaviors are important. Problems where the effectiveness of the intervention in realistic settings may be different from settings of clinical trials because of differences in behaviors.
- Problems where errors in tests and treatments are important.
- Problems where timing is important, such as screening, timing follow-up visits, or decisions about how long to keep a patient on a treatment before modifying it or switching to another.

Diseases in which important events unfold unevenly over time, such as chronic diseases that have occasional acute events:

- Problems that span multiple diseases and/or interventions, such as setting priorities.
- Problems that involve performance, such as performance measures and pay for performance.

For problems like these we have found it necessary to design a new type of model that combines differential equations, OOP, and features. We believe this approach is new, at least in health. A PubMed search for text words "differential equations" and "object oriented" begins with our 2003 article. Since then three additional articles have been published applying differential equations and object oriented methods to cell simulation. The offer then is that if others find themselves facing similar constraints with other types of models, they might find it useful to try this approach.

References

1 Gersh W, Eddy DM, Dong E. Cardiac arrhythmia classification: a heart-beat interval-markov chain approach. Comput Biomed Res 1970;4:385–92.
2 Frick MH, Elo O, Haapa K, et al. Helsinki Heart Study: primary-prevention trial with gemfibrozil in middle-aged men with dyslipidemia: safety of treatment, changes in risk factors, and incidence of coronary heart disease. N Engl J Med 1987;317: 1237–45.
3 The lipid research clinics coronary primary prevention trial results. I. Reduction in incidence of coronary heart disease. JAMA 1984;251:351–64.
4 Shepherd J, Cobbe SM, Ford I, et al. Prevention of coronary heart disease with Pravastatin in men with hypercholesterolemia. N Engl J Med 1995;333:1301–7.
5 Heart Protection Study Collaborative Group. MRC/BHF Heart Protection Study of antioxidant vitamin supplementation in 20,536 high-risk individuals: a randomized placebo-controlled trial. Lancet 2002; 360:23–33.
6 The Heart Outcomes Prevention Evaluation Study Investigators. Effects of Ramipril on cardiovascular and microvascular outcomes in people with diabetes mellitus: results of the HOPE study and MICRO-HOPE substudy. Lancet 2000;355:253–9.
7 LIPID Study Group. Prevention of cardiovascular events and death with pravastatin in patients with coronary heart disease and a broad range of initial cholesterol levels. N Engl J Med 1998;339:1349–57.
8 Sacks FM, Pfeffer MA, Moye LA, et al. The effect of Pravastatin on coronary events after myocardial infarction in patients with average cholesterol levels. N Engl J Med 1996;335:1001–9.
9 Scandinavian Simvastatin Survival Study Group. Randomized trial of cholesterol in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). Lancet 1994;344: 1383–9.
10 Eddy DM. A methodology for evaluating breast cancer screening programs. Proceedings of the International Conference on Cybernetics and Society, November, 1976.
11 Eddy DM. Guidelines for the cancer-related checkup: recommendations and rationale. CA Cancer J Clin 1980;30:193–240.
12 Gold MR, Siegel JE, Russell LB, Weinstein MC. Cost-Effectiveness in Health and Medicine. New York: Oxford University Press, 1996.
13 Brody JE. Cancer society reports it finds some detection tests unneeded. New York Times, March 21, 1980, p. 1.
14 Eddy DM. The frequency of cervical cancer screening. Comparison of a mathematical model with empirical data. Cancer 1987;60:1117–22.
15 Eddy DM. Screening for Cancer: Theory, Analysis and Design. Englewood Cliffs, NJ: Prentice Hall Inc, 1980.
16 Lanchester Prize, 1980. Available from: http://www.informs.org/Prizes/LanchesterPrize.html [Accessed November 14, 2005].
17 Schlessinger L, Eddy DM. Archimedes: a new model for simulating health care systems: the mathematical formulation. J Biomed Inform 2002;35:37–50.
18 Eddy DM, Schlessinger L. Archimedes. a trial-validated model of diabetes. Diabetes Care 2003;26: 3093–101.
19 Eddy DM, Schlessinger L. Validation of the Archimedes diabetes model. Diabetes Care 2003;26: 3102–10.
20 UK Prospective Diabetes Study (UKPDS) Group. Intensive blood-glucose control with sulphonylureas or insulin compared with conventional treatment and risk of complications in patients with type 2 diabetes (UKPDS 33). Lancet 1998;352:837–52.
21 The Diabetes Control and Complications Trial Research Group. The effect of intensive treatment of diabetes on the development and progression of long-term complications in insulin-dependent diabetes mellitus. N Engl J Med 1993;329:977–86.
22 Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with life-

style intervention or metformin. N Engl J Med 2002;356:393–402.
23 Colhoun HM, Betteridge DJ, Durrington PN, et al. Primary prevention of cardiovascular disease with atorvastatin in type 2 diabetes in the Collaborative Atorvastatin Diabetes Study (CARDS): multicenter randomized placebo-controlled trial. Lancet 2004; 364:685–96.